Automatic Exercise Generation Based on Corpora and Natural Language Processing Techniques

Candidate: Itziar Aldabe
Supervisor: Montse Maritxalar

Lengoaia eta Sistema Informatikoak Saila
Euskal Herriko Unibertsitatea / University of the Basque Country

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Information and communication technologies (ICT)

- The introduction of ICT into educational areas
  - ICT competence
  - Classroom
  - Distance learning
- Examples
  - Eskola 2.0 (Basque Gov.)   Previously
  - Ikasys (Ikastolen Elk.) $\Rightarrow$ established
  - Ikasbil (HABE)   domains and content

Static knowledge $\Rightarrow$ Dynamic knowledge

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Objectives

- The study of different NLP tools and resources in order to generate exercises in an automatic way

- To offer experts a helpful tool with which to create didactic material

- The obtaining of pedagogically useful exercises

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Objectives

- The study of different NLP tools and resources in order to generate exercises in an automatic way
  - Student's knowledge of English (Hoshino and Nakagawa 2007; Pino *et al.*, 2008; Sumita *et al.*, 2005)
  - Specific domain (Mitkov *et al.*, 2009)
  - Other languages: Bulgarian (Nikolova 2009), Irish (Foster *et al.*, 2011)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Question, stem and distractor

Example

...1... hori behar bezala ez kanporatzeagatik, gerta daiteke birikek duten baino leku gehiago behar izatea aireak, eta presio horrek biriketako albeoloei eztanda eginaraztea. Albeoloak lehertzean sortzen den odoljarioak ...2... ditu, bai urpekariak, bai saguzarrak

1	a. Aire	b. Haize	c. Oxigeno	d. Gas
2	a. gaixotzen	b. mareatzen	c. akabatzen	d. desorientatzen

Question or item

A stem that requires an answer (key)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Question, stem and distractor

Stem

...1... hori behar bezala ez kanporatzeagatik, gerta daiteke birikek duten baino leku gehiago behar izatea aireak, eta presio horrek biriketako albeoloei eztanda eginaraztea. *Albeoloak lehertzean sortzen den odoljarioak ...2... ditu, bai urpekariak, bai saguzarrak*

| 1 | a. Aire | b. Haize | c. Oxigeno | d. Gas |
| 2 | a. gaixotzen | b. mareatzen | c. akabatzen | d. desorientatzen |

Stem

A problem to be solved, a question, or an incomplete statement

Question, stem and distractor

Correct answer

...1... hori behar bezala ez kanporatzeagatik, gerta daiteke birikek duten baino leku gehiago behar izatea aireak, eta presio horrek biriketako albeoloei eztanda eginaraztea. Albeoloak lehertzean sortzen den odoljarioak ...2... ditu, bai urpekariak, bai saguzarrak

1   **a. Aire**   b. Haize   c. Oxigeno   d. Gas
2   a. gaixotzen   b. mareatzen   c. akabatzen   d. desorientatzen

Correct answer

The correct answer to the stem is the key of the item

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Question, stem and distractor

Distractors

...1... hori behar bezala ez kanporatzeagatik, gerta daiteke birikek duten baino leku gehiago behar izatea aireak, eta presio horrek biriketako albeoloei eztanda eginaraztea. Albeoloak lehertzean sortzen den odoljarioak ...2... ditu, bai urpekariak, bai saguzarrak

1   a. Aire   **b. Haize**   **c. Oxigeno**   **d. Gas**
2   a. gaixotzen   b. mareatzen   c. akabatzen   d. desorientatzen

Distractor

An incorrect choice among multiple-choice answers on a test

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Question, stem and distractor

Example

...1... hori behar bezala ez kanporatzeagatik, gerta daiteke birikek duten baino leku gehiago behar izatea aireak, eta presio horrek biriketako albeoloei eztanda eginaraztea. Albeoloak lehertzean sortzen den odoljarioak ...2... ditu, bai urpekariak, bai saguzarrak

1	a. Aire	b. Haize	c. Oxigeno	d. Gas
2	a. gaixotzen	b. mareatzen	c. akabatzen	d. desorientatzen

Topic

The topic is the concept that students have to work with and is part of their curriculum

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Contributions

- **ArikIturri**
  - A system which generates items automatically from corpora, using NLP techniques
  - Different types of item
  - Modular, multilingual and independent of any application

- **Question model**
  - Representation of the items generated by ArikIturri
  - General and flexible

- **Experiments focusing on different scenarios**
  - Basque language learning
  - English language learning
  - Learning in the science domain

- **Evaluation of the items**
  - Experts' evaluation
  - Evaluation with students

Tasks

Generation of the items

- Components
    - Identification of the stems
    - Generation of the distractors

- Multilingualism feature

- Grammatical and semantic information

Evaluation of the items

- Experts' evaluation

- Evaluation with students

Introduction
Architecture
Model
Experiments
Conclusions and future work

Motivation

Outline

1 **Introduction**

2 **Architecture**

3 **Model**

4 **Experiments**

5 **Conclusions and future work**

ArikIturri

- ArikIturri aims to be a source of items with which to create a bank of useful exercises to be used within the cognitive domain
  - Input texts
  - Two kinds of language resource: NLP tools and linguistic information
- Multilingual: Basque and English
- Modular system: reusability and portability

ArikIturri

- Types of question
  - Fill-in-the-blank
  - Word formation
  - Multiple choice
  - Error correction
  - Short answer
- Topic
  - Basque grammar: determiners, declension and verbs
  - English vocabulary: verbs
  - Basque science: vocabulary

ArikIturri: architecture

Architecture
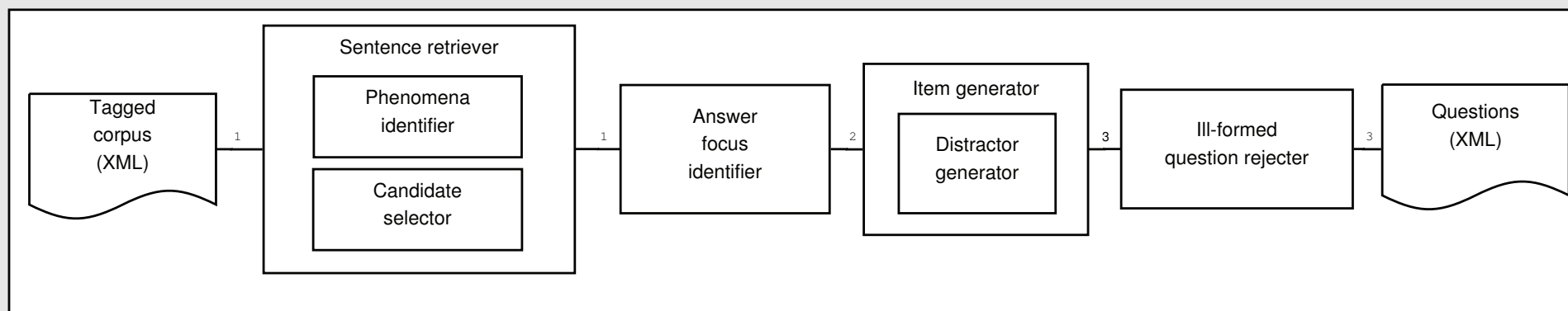
- Input: morphologically and syntactically analysed texts (XML)
- Output: items represented in XML
- Modular: results in the intermediate steps

NLP tools and language resources

Architecture
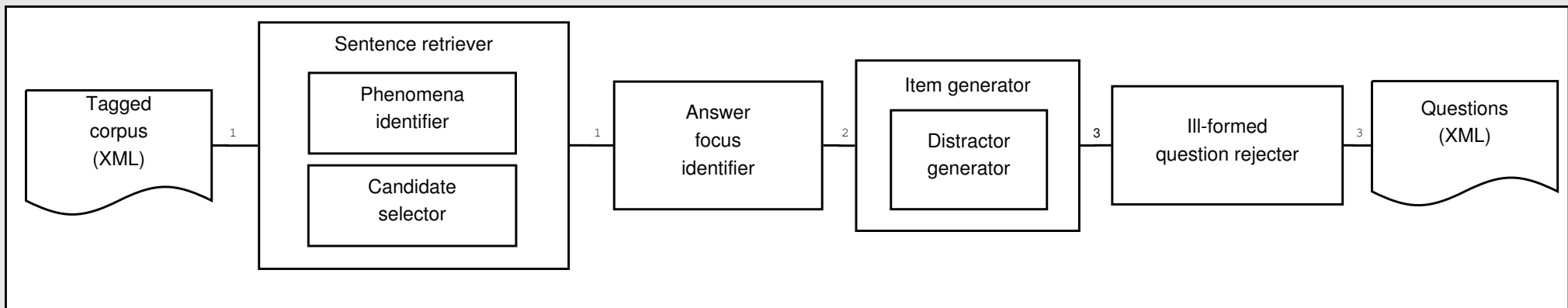
NLP tools and language resources

Sentence retriever and Item generator
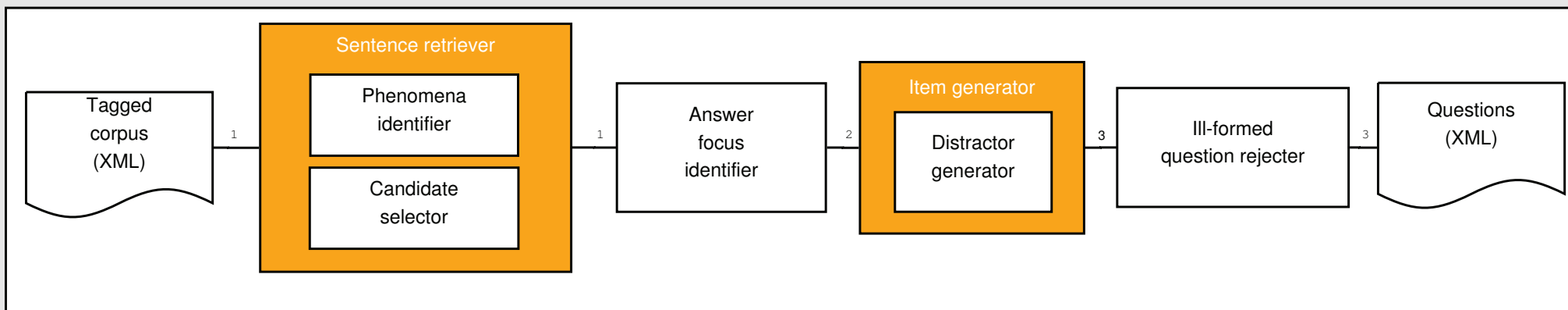
NLP tools and language resources

Input corpora

Senten

Corpus

Phe
id

Ca
se

ZT corpus: Stem generation

Basque language learning: Declension and verb test

Basque learner: Determiner test

BNC: MCQs for English verbs

5 articles from zientzia.net: MCQs for Basque science vocabulary

NLP tools and language resources

Analysers

Corpus

Sentence

Pheno
ident

Candi
sele

Basque chunker: Ixati
Connexor machinese syntax

em

D
g

Tagged
Corpus
(XML)

cter

3

Questions
(XML)
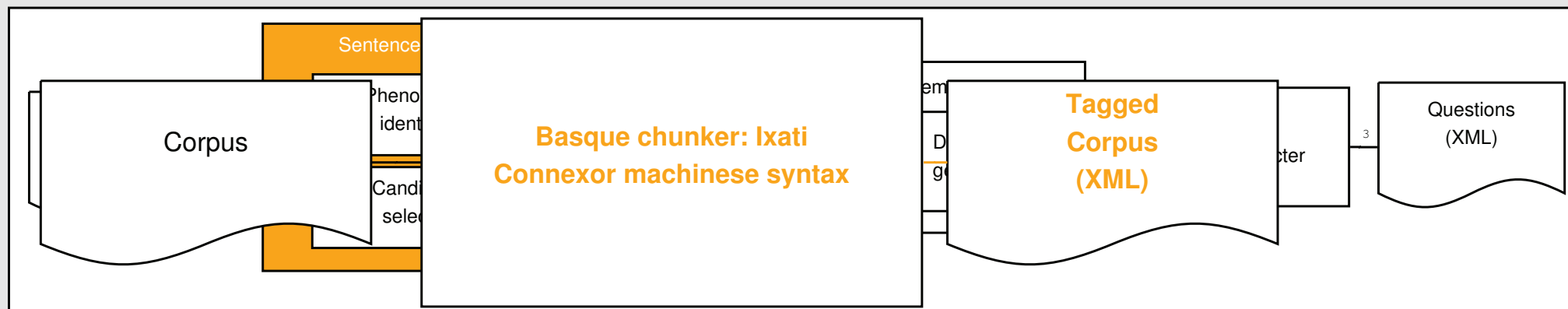
NLP tools and language resources

Item generator
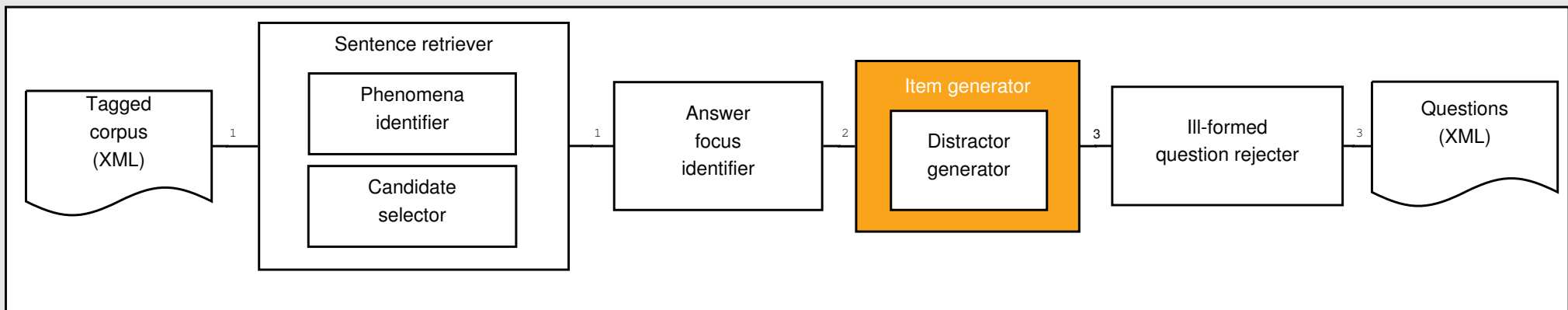
NLP tools and language resources

Corpora

* Corpora as a measure of similarity
    - ZT corpus
    - BNC
* Corpora to search for occurrences
    - Euskaldunon egunkaria
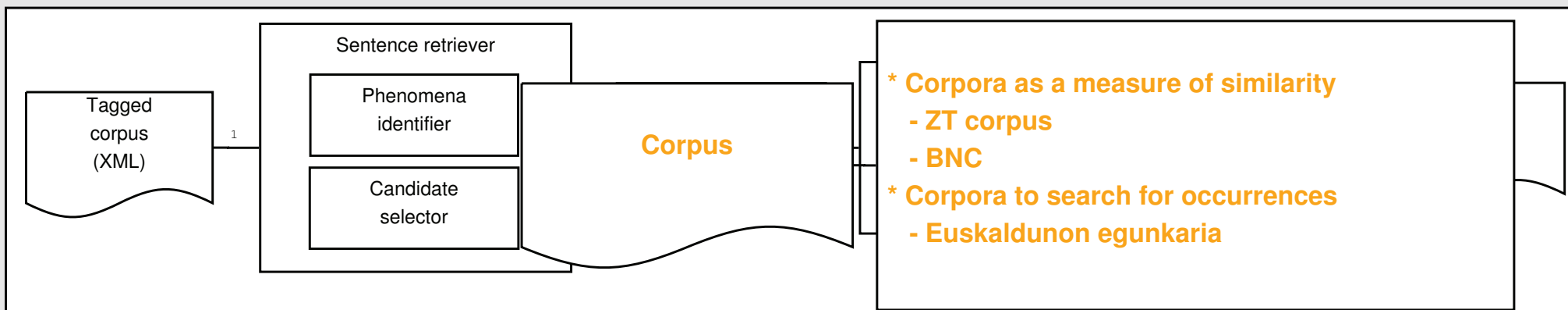
NLP tools and language resources

Ontologies and dictionaries

Tagged corpus (XML)

1

Sentence retriever

Phenomena identifier

Candidate selector

1

Answer focus identifier

2

* Ontologies and dictionaries
 - Monolingual dictionary
 - Encyclopedic dictionary of science
 - WordNet and MCR
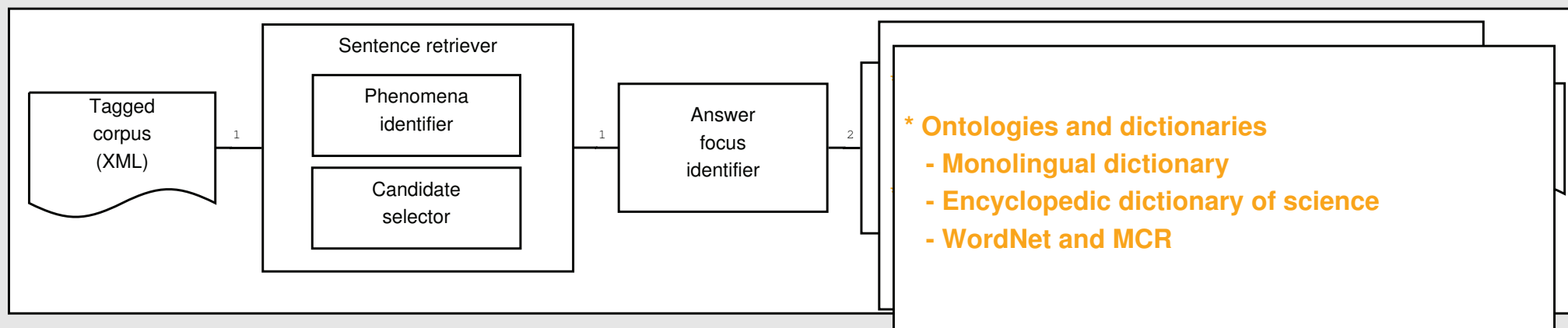
NLP tools and language resources

Basque generator
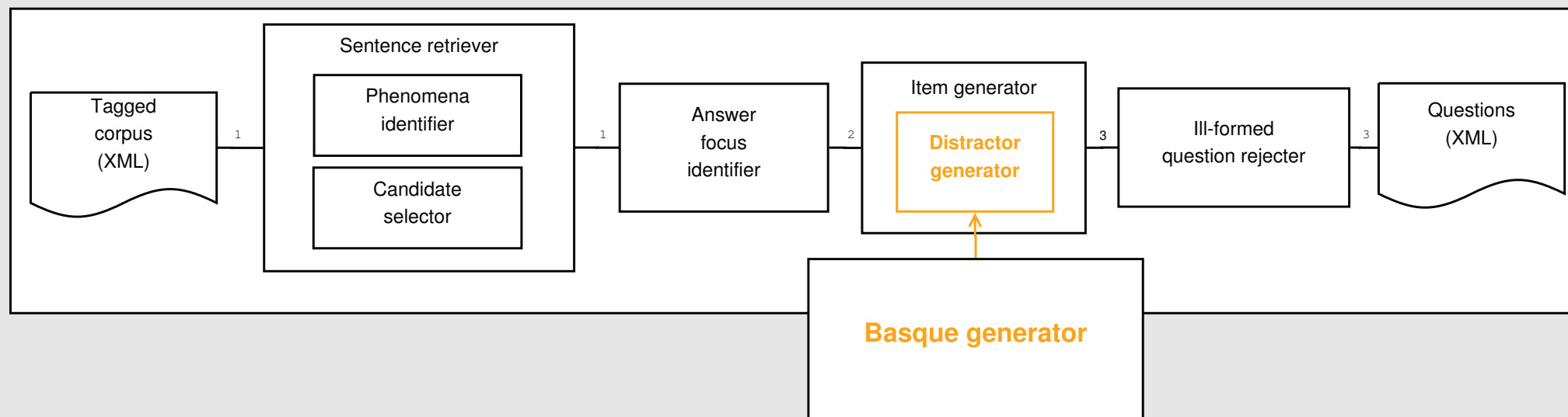
NLP tools and language resources

Post-editing environment
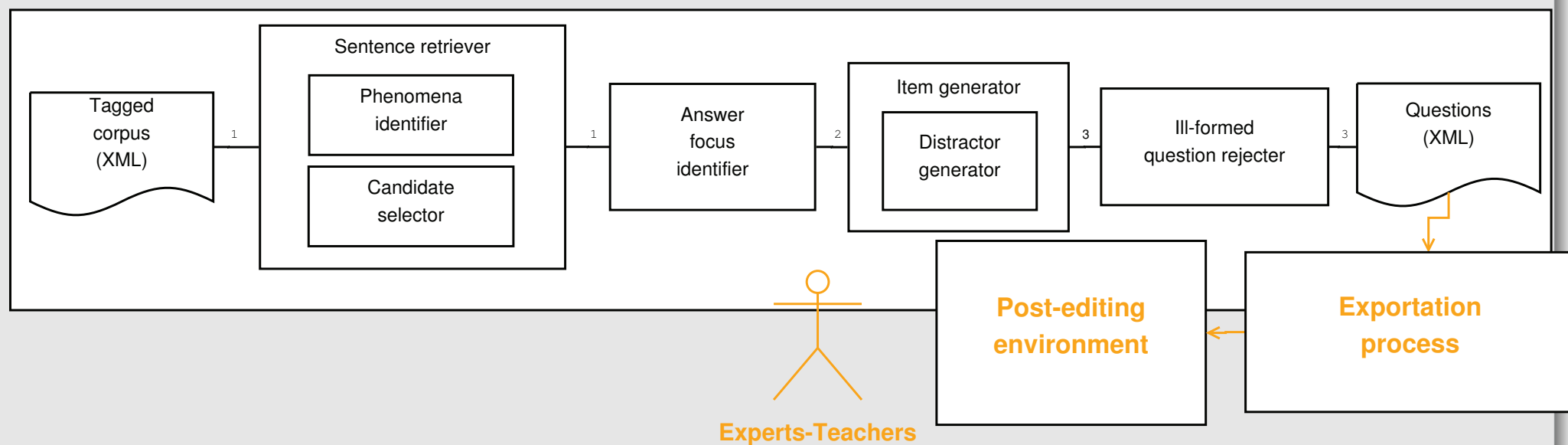
Outline

1. Introduction

2. Architecture

3. Model

4. Experiments

5. Conclusions and future work

Question

Question

- A question is a sentence or clause in which the topic which the student has to focus on appears and is represented as part of the text as a whole
- A question contains all of the information relating to a particular item
  - stem, key, distractors, information relating to the generation process

Model

The question model comprises all of the components of the items, and depending on the question type, the question will have certain specific components

ArikIturri: question model

- Structured representation of the items, topical information and heuristics
  - Experts can consult the heuristics in order to better understand the generation process ⇒ Feedback
  - Information about the generation process provides hints to guide students in a technology-enhanced learning system
  - Diagnosis

- The items generated by ArikIturri should be represented in a standard way to ensure their durability, portability and reusability

- A two step process
  - Own specification
  - QTI extension

Question

- Each item:
  - Corresponds to a particular source sentence
  - The sentences are represented at the chunk level
- A question, conceptually, has three main components: the topic, the answer focus and the context
  - The answer focus is the chunk of the sentence in which the topic appears
  - The rest of the chunks of the sentence are collected into the context element

Example: fill-in-the-blank

Source sentence

Sintomak hauek dira: aldarte txarra, estresa eta antsietatea

Fill-in-the-blank

Sintomak hauek ....: aldarte txarra, estresa eta antsietatea

- Fill-in-the-blanks have at least one blank that students need to fill in
- An item designed to deal with Basque verbs

Example: fill-in-the-blank

Answer focus

```xml
<answer_focus posQ="2" posS="2" change="false" blank="true">
<head>
<answer>
<word pos="0">dira</word>
<topic_info>
<linguistic_info>
mode(indicative);tense(present);subj(they)
</linguistic_info>
</topic_info>
<analysis pos="0">"izan"  ADT PNT A1 NOR NR_HAIEK @+JADNAG %ADIKAT</↵
    analysis>
</answer>
</head>
</answer_focus>
```

Example: fill-in-the-blank

Answer focus

```
<answer_focus posQ="2" posS="2" change="false" blank="true">
<head>
<answer>
<word pos="0">dira</word>
<topic_info>
<linguistic_info>
mode(indicative);tense(present);subj(they)
</linguistic_info>
</topic_info>
<analysis pos="0">"izan" ADT PNT A1 NOR NR_HAIEK @+JADNAG %ADIKAT</↩
    analysis>
</answer>
</head>
</answer_focus>
```

Example: fill-in-the-blank

Answer focus

```
<answer_focus posQ="2" posS="2" change="false" blank="true">
<head>
<answer>
<word pos="0">dira</word>
<topic_info>
<linguistic_info>
mode(indicative);tense(present);subj(they)
</linguistic_info>
</topic_info>
<analysis pos="0">izan    ADT PNT A1 NOR NR_HAIEK @+JADNAG %ADIKAT</←
    analysis>
</answer>
</head>
</answer_focus>
```

Example: error correction

Error correction example

Hainbat ariketaren bidez <u>gure gorputzaren blokeoarekin</u> askatu behar dugu

- A marked error correction item designed to deal with declension
- New element: distractor

Example: error correction

Distractor

```
<distractor>
<topicGroup> B2 </topicGroup>
<topicGroup> C1 </topicGroup>
<word pos="2">blokeoarekin</word>
<heuristic>
<type> declension </type>
<function> replacement(basque, abs, soz) </function>
<input> blokeoa </input>
<output> blokeoarekin </output>
</heuristic>
<analysis pos="2">
("blokeo"  IZE ARR DEK SOZ NUMS MUGM @ADLG %SIB)
</analysis>
</distractor>
```

Example: error correction

Distractor

```
<distractor>
<topicGroup> B2 </topicGroup>
<topicGroup> C1 </topicGroup>
<word pos="2">blokeoarekin</word>
<heuristic>
<type> declension </type>
<function> replacement(basque, abs, soz) </function>
<input> blokeoa </input>
<output> blokeoarekin </output>
</heuristic>
<analysis pos="2">
("blokeo"  IZE ARR DEK SOZ NUMS MUGM @ADLG %SIB)
</analysis>
</distractor>
```

Example: error correction

Distractor

```
<distractor>
<topicGroup> B2 </topicGroup>
<topicGroup> C1 </topicGroup>
<word pos="2">blokeoarekin</word>
<heuristic>
<type> declension </type>
<function> replacement(basque, abs, soz) </function>
<input> blokeoa </input>
<output> blokeoarekin </output>
</heuristic>
<analysis pos="2">
("blokeo"   IZE ARR DEK SOZ NUMS MUGM @ADLG %SIB)
</analysis>
</distractor>
```

Procedure

- The option of representing the question types implemented by ArikIturri in QTI
  - The model does not offer the option of representing the information relating to the source sentence, topic and generation process
- An extension point is needed
  - The extension points offered in the QTI model are not appropriate to represent the information relating to the generation process

QTI extension

Two modifications

- An attribute to represent information regarding the **topic**
- The way to represent the **head** of the answer focus defined in our model

QTI extension

- Modification of the *assessmentItem* class
  - An attribute (topic)
  - One element (head)

QTI extension: assessmentItem

```xml
<xsd:attributeGroup name=" assessmentItem.AttrGroup">
<xsd:attribute name=" identifier" type=" string.Type" use=" required"/>
...
<xsd:attribute name=" topic" type=" string.Type" use=" optional"/>
</xsd:attributeGroup>

<xsd:group name=" assessmentItem.ContentGroup">
<xsd:sequence>
<xsd:element ref=" responseDeclaration" minOccurs="0" maxOccurs="↵
    unbounded"/>
<xsd:element ref=" outcomeDeclaration" minOccurs="0" maxOccurs="↵
    unbounded"/>
...
<xsd:element ref=" head" minOccurs="0" maxOccurs="1"/>
</xsd:sequence>
</xsd:group>

<xsd:complexType name=" assessmentItem.Type" mixed=" false">
<xsd:group ref=" assessmentItem.ContentGroup"/>
<xsd:attributeGroup ref=" assessmentItem.AttrGroup"/>
</xsd:complexType>

<xsd:element name=" assessmentItem" type=" assessmentItem.Type"/>
```

QTI extension: head Element

head.Type

```
<xsd:complexType name="head.Type">
  <xsd:sequence>
    <xsd:element name="answer" type="answer.Type"/>
    <xsd:element name="distractor" type="distractor.Type" minOccurs="0" ↩
        maxOccurs="unbounded"/>
    <xsd:element name="headComponent" type="headComponent.Type" minOccurs="↩
        0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>
```

QTI extension: head Element

answer.Type

```xml
<xsd:complexType name="answer.Type">
  <xsd:sequence>
    <xsd:element name="word" type="word.Type" maxOccurs="unbounded"/>
    <xsd:element name="topic_info" type="topic_info.Type" />
    <xsd:element name="analysis" type="analysis.Type" maxOccurs="unbounded"←
        />
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="topic_info.Type">
  <xsd:sequence>
    <xsd:choice>
      <xsd:element name="linguistic_info" type="string.Type"/>
      <xsd:element name="lemma" type="string.Type"/>
    </xsd:choice>
    <xsd:element name="function" type="string.Type" minOccurs="0"/>
    <xsd:attribute name="artificial" type="boolean.Type" use="optional"/>
    <xsd:any minOccurs="0"/>
  </xsd:sequence>
</xsd:element>
```

QTI extension: head Element

distractor.Type

```xml
<xsd:complexType name="distractor.Type">
  <xsd:sequence>
    <xsd:element name="topicGroup" type="topicGroup.Type" use="required" ↵
          minOccurs="0" maxOccurs="unbounded" />
    <xsd:element name="word" type="word.Type" maxOccurs="unbounded" />
    <xsd:element name="heuristic" type="heuristic.Type" />
    <xsd:element name="order" type="order.Type" minOccurs="0" />
    <xsd:element name="analysis" type="analysis.Type" maxOccurs="unbounded" ↵
          />
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="heuristic.Type">
  <xsd:sequence>
    <xsd:element name="type" type="string.Type" />
    <xsd:element name="function" type="string.Type" />
    <xsd:element name="input" type="string.Type" />
    <xsd:element name="output" type="string.Type" />
  </xsd:sequence>
</xsd:complexType>
```

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Outline

1. Introduction

2. Architecture

3. Model

4. Experiments

5. Conclusions and future work

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experiments

1. Question generation
2. Declension and verb tests
3. Determiner tests
4. MCQs for English verb tests
5. MCQs for Basque science tests

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Task

- Stem generation task: transformation of the source sentence
- Question generation challenge: Text-to-question task (Rus and Graesser, 2009)
    - Target selection $\Rightarrow$ *Phenomena identifier*
    - Question type selection $\Rightarrow$ *Item generator*
    - Question construction $\Rightarrow$ *Item generator*
- Experiment
    - Interrogative stems (questions) regarding numerical entities
    - ZT corpus; Basque language
    - Independent task

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Question generation — Example

Sentence from ZT corpus

Joan den abenduan argitaratu zuen txostena, eta **otsailaren 25a arte***, nahi duenak iritzia emateko aukera du.*

The report was published last December, and, those who want to do so have the opportunity to express their views **until February 25**th.

Question generated from the source sentence

NOIZ ARTE du nahi duenak iritzia emateko aukera?

UNTIL WHEN do those who want to do so have the opportunity to express their views?

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Target selection

- Clause identification
- Numerical entity identification
- Candidate selection

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Target selection

- Clause identification
  - Coordinated clauses from the source sentence
  - A combination of rule-based grammar with machine learning techniques (Alegria *et al.*, 2008)

- Numerical entity identification

- Candidate selection

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Target selection

- Clause identification
- Numerical entity identification
  - Numerical Entity Recogniser and Classifier for Basque (NuERCB) (Soraluze *et al.*, 2011)
  - NuERCB decides whether the numbers express a date or time, are associated with units of measurement, or refer to common nouns
- Candidate selection

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Target selection

- Clause identification

- Numerical entity identification

- Candidate selection
  - One main verb
  - The shortest candidate clauses

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Question type selection

- The QG system is responsible for identifying the corresponding wh-word
- The Basque wh-words make different word formation
  - Patterns to recognise the numerical entities and the morphosyntactic information

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Question construction

- Modifications
  - Main verb in the first singular or plural $\Rightarrow$ Third person
  - Deletion of linking words to connect clauses
  - More than one numerical entity $\Rightarrow$ The closest to the verb
- Construction; Transformation rules
  - Beginning of the question: the wh-word followed by the rest of the words of the chunk
  - The main verb
  - The rest of the chunks to the right of the verb
  - The chunks that appear on the left

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation

Syntactic correctness and fluency

Rank	Description	Changes	Result
1	The question is grammatically correct and idiomatic/natural	No changes	39.34%
2	The question is grammatically correct but does not read fluently	Minor changes	22.95%
3	There are some grammatical errors in the question	Major changes	9.83%
4	The question is grammatically unacceptable	Discard	27.86%

Table: Scoring for syntactic correctness and fluency

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation

Question type evaluation

- 85.24% of wh-words correspond to the source sentence
- 88.52% of the generated questions related to the source sentence

System's performance

- Precision: 84.25%
- Recall: 78.26%

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experiments

1. Question generation

2. Declension and verb tests

3. Determiner tests

4. MCQs for English verb tests

5. MCQs for Basque science tests

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Aim

- The automatic generation of items in a Basque language learning scenario
- Aim to test the language level of students: declension and verb
- Question types: error correction and multiple-choice question (MCQ)
- Definition of the heuristics
  - Errors made by learners
  - Experts in the field
  - An attempt to automatise this process

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Declension and verb tests — Example

Error correction example — Declension

Badaude <u>beldurrari</u> zerikusia duten barreak ere

(There are also some kinds of laughter which have to do <u>to</u> fear)

Multiple-choice example — Verb

Sintomak honako hauek .... : aldarte txarra, estresa eta antsietatea

(The symptoms .... : bad-mood, stress and anxiety)

a) **dira** (are) (**key**)
b) da (is) (distractor)
c) daude (are) (distractor)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Handmade heuristics

Topics and level

- Five inflection cases and two verb forms: the sociative, inessive, absolutive, dative and ergative cases and present and past indicative verb tenses
- Input corpus: the language learning corpus
  - The source data came from the high language level corpus; C1 level students

Heuristics

- Declension cases: the incorrect use of declension cases or finiteness
- Verb tenses: change the different persons of the verb that belong to different auxiliary paradigms

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation - Experimental settings

Experimental settings

- Error correction: the correct use of the inflection cases
- MCQ
  - The finiteness of the inflection cases (3 distractors)
  - Verb forms (2 distractors)
- 1700 error correction and MCQs

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Correctness of the questions

1st manual evaluation

- One expert
- Accepted questions: 83.36% error correction and 82.71% MCQs

2nd manual evaluation

- Agreement between the experts
  - Total agreement: 0.41 (moderated)
  - MCQs for verb tenses: 0.61 (substantial)
  - Error correction for declension cases: 0.28 (fair)

Discussion

- The number of distractors changes the acceptance rate
- The topic may also have an influence on the results

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation and generation criteria of experts

Motivation

- Experts experience in order to improve the generation process
  - Reasons for accepting, discarding or modifying a stem or a distractor

Appropriateness of the stem

- ArikIturri no longer allowed the generation of items with a blank in the first position
- Reasons for discarding stems
  - Inappropriateness of the sentence length
  - The need for a larger context
  - The stem is too difficult for learners

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation and generation criteria of experts

Motivation

- Experts experience in order to improve the generation process
  - Reasons for accepting, discarding or modifying a stem or a distractor

New ways of generating distractors

- To change the paradigm of one candidate distractor per item on average
- The addition of some new elements. For example: the subordinating verbal prefix "bait-"
- To change the root or tense of the verb
- The deletion of some elements

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation and generation criteria of experts

Motivation

- Experts experience in order to improve the generation process
  - Reasons for accepting, discarding or modifying a stem or a distractor

Discussion

- The help of experts is really useful in order to improve the generation and evaluation of the items
- This way of acquiring knowledge to be integrated into ArikIturri is expensive
- Alternative ways to generate distractors: automatically extracted patterns

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experiments

1. Question generation
2. Declension and verb tests
3. Determiner tests
4. MCQs for English verb tests
5. MCQs for Basque science tests

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Motivation

- A learner corpus contains a collection of errors made by learners $\Rightarrow$ An alternative way of defining heuristics

- Manually tagged determiner errors $\Rightarrow$ The repetition of determiners within the same phrase

- Uria (2009)
  - Rules based on CG formalism for the automatic treatment of determiner errors
  - 58 rules to deal with the repetition of determiners

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Determiner test — Example

MCQ example — Determiners

Nire bizitzaren .... orain dela 5 urte gertatu zen.

(.... in my life took place 5 years ago.)

a) **egun zoriontsu bat** (one happy day) (**correct answer**)
b) eguna zoriontsua bat (a one happy the day) (distractor – learner corpus)
c) eguna zoriontsu bat (one happy the day) (distractor)
d) egun zoriontsuak bat (the one happy day) (distractor)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Heuristics

- We first grouped the rules according to the erroneous phrase they detected ⇒ Different heuristics for each sub-group
- The basis for defining the heuristics
  - The error types of the rest of sub-groups
  - The rest of the determiner error types
  - Some other error types which are closely related ⇒ For example, the incorrect use of declension cases
- This type of item also needs a correct answer ⇒ Automatically generated

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation

Experimental settings

- The source text: one text written by a low-level learner who was asked to write a description of the happiest day in her/his life
- Candidate distractors
    - Repetition of the determiner (R_DET)
    - Wrong order of the determiner (WO_DET)
    - Use of the wrong determiner (W_DET)
    - A combination of two error types (WO_DET and W_DET)
    - Errors relating to declension cases
- 30 low-language levels learners; mother tongue Spanish

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Evaluation

Results

- Not very difficult test
    - 60% of the students: 3 mistakes or fewer
- There are some students that made at least 4 mistakes
    - There are NO examples of this behaviour in the learner corpus

Hypotheses

- Offering this type of test leads to the emergence of errors that students do not produce in a writing task
- Learners tend to write in a simpler way in order to avoid making errors

Discussion

- Learner corpus as a source to define the heuristics

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experiments

1. Question generation
2. Declension and verb tests
3. Determiner tests
4. MCQs for English verb tests
5. MCQs for Basque science tests

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Aim

Aim

- Multilinguality
- Automatic selection of the source sentence
- Automatic generation of distractors

Experiment

- Isolated MCQs of English vocabulary $\Rightarrow$ Verbs from the Academic Word List (AWL)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Aim

Aim

- Multilinguality
- Automatic selection of the source sentence
- Automatic generation of distractors

Experiment

- Isolated MCQs of English vocabulary ⇒ Verbs from the Academic Word List (AWL)
- The ability of the system to automatically generate verbs as distractors which are **semantically** similar to the key
- All of the verbs which are part of the AWL were considered as candidate distractors
- 200 MCQs chosen at random

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

MCQs for English verb test — Example

MCQ to test the verb "respond" from the AWL

Certainly, people .... to whether they perceive the world as threatening or reassuring.

a) **respond**   (**key**)
b) identify   (distractor)
c) function   (distractor)
d) interpret   (distractor)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experimental settings

Baseline system

- The verbs which are most similar to the key were selected as candidate distractors, employing the **information radius** (corpus cooccurrences;ranked)
- Matching in terms of transitiveness/intransitiveness, tense and person

Gold-standard

- 94.07% of the distractors which were part of the accepted questions as valid

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Language model from the Web 1T 5-gram dataset

Aim

- We studied whether the context could play a beneficial role in the selection of distractors
- Language model from the Web 1T 5-gram dataset

Heuristic

- Candidate distractor probability
  - Lower than a sixth of the correct 3-gram probability
  - Greater than zero
  - Empirically established

Results

- Precision: 94.30%
- Recall: 37.84%
- The heuristic offers a way to select different distractors, taking into account the context of the candidate sentence

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Discussion

- Multilingual system

- ArikIturri adds semantic information to grammatical information when generating distractors

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experiments

1. Question generation
2. Declension and verb tests
3. Determiner tests
4. MCQs for English verb tests
5. MCQs for Basque science test

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Introduction

- The aim is to propose distractors that correspond to the vocabulary studied as part of students' curricula

- Basque scientific vocabulary within the science and technology domain

- Semantic relatedness measures in order to obtain the distractors

- Real scenario

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

MCQs for Basque science test — Example

MCQs embedded in the source text — Basque

Espazioan itzalkin erraldoi bat ezartzeak, bestalde, Lurrari...1... egingo lioke, poluitu gabe. Siliziozko milioika disko ...2... bidaltzea da ikertzaileen ideia. Paketetan jaurtiko lirateke, eta, behin diskoak zabalduta, itzalkin-itxurako egitura handi bat osatuko lukete. Hori bai,...3... handiegiak izango lituzke.

1	a. babes	b. aterki	c. defentsa	d. **itzala**
2	a. unibertsora	b. izarrera	c. galaxiara	d. **espaziora**
3	a. **kostu**	b. prezio	c. eragozpen	d. zailtasun

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Steps

1. Selection of the texts   ⇒ Experts
2. Marking the blanks   ⇒ Experts
3. Generation of distractors
4. Choosing the distractors   ⇒ Experts
5. Evaluation with students   ⇒ In-class experiment
6. Item Analysis

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Methods

- Generation approaches
  - Corpus-based approaches
    - LSA-based
    - LSA-based & ontology & morphology
    - LSA-based & specialised dictionary
  - Graph-based approach
  - Combination of corpus- and graph-based approaches
- Our system employs the context in which the key appears to compute the similarities

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA

- The system uses context-words to compute the similarity deploying Latent Semantic Analysis (LSA)

- LSA is a theory and method for extracting and representing the meaning of words

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA

- The baseline system provides the whole sentence where the key appears
- Candidate distractors: the first words of the output which are not part of the sentence and match the same PoS
- A generation process is applied

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA & ontology & morphology

Example

Istripua izan ondoren, ... sendatu ninduen
a) medikuak (*the doctor*)
b) ospitalak (*the hospital*)

- Both words are related and belong to the same specific domain

Students could discard *ospitalak* as the answer, knowing that the correct answer has to be a person in the given sentence

The system tries to avoid this kind of guessing by means of semantic information

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA & ontology & morphology

- Two semantic resources:
  - Semantic features of common nouns obtained with a semiautomatic method
    - Semantic features: animate, human, concrete etc.
  - The Basque WordNet and the Multilingual Central Repository (MCR)
    - The properties of the Top Concept Ontology, the WordNet Domains and Suggested Upper Merged Ontology (SUMO)

- It proposes only the candidate distractors that share at least one semantic characteristic with the key

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA & ontology & morphology

Two words may share a semantic characteristic but differ in their inflection features

ospitaleko, medikuko

The system tries to avoid candidate distractors whose highly improbable inflected form would give to the students a clue to reject them

Once distractors are generated, the system searches for any occurrence of the new inflected word in a corpus

- If there is any occurrence ⇒ Candidate distractor

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA & specialised dictionary

- The method combines the information offered by the LSA model and the entries of an encyclopaedic dictionary of Science and Technology for Basque
- Dictionary:
    - 23,000 basic concepts
    - 50 different subjects

- The system searches in the dictionary the lemmas of the key and the distractors
- Preference
    1. the candidate distractors which share the subject with the key
    2. the candidates which share any semantic characteristics with the key

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Graph-based approach

- Graph-based approach first introduced by Agirre and Soroa (2009)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Graph-based approach

1. A list of candidate distractors is obtained from WordNet
   - Monosemous key $\Rightarrow$ All of its siblings
   - More than one meaning $\Rightarrow$ Obtain its most likely sense and then its siblings
   - No siblings $\Rightarrow$ The hyponyms of the key

2. The personalised PageRank vector is obtained for the given context and the key

3. The personalised PageRank vectors are obtained for 20 candidate distractors in the given context

4. The similarities among the vectors computed by the dot product are measured and a list of reordered candidate distractors is obtained

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

LSA & graph-based

- Similarity is computed in two rounds
  1. The system obtains a ranked list of candidate distractors based on the LSA method
  2. The Personalised PageRank vector is obtained for the stem and the key
  3. The Personalised PageRank vectors are obtained for the stem and 20 candidate distractors
  4. The similarities among vectors computed by the dot product are measured and a reordering of the candidate distractors is obtained

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Steps

1. Selection of the texts $\Rightarrow$ Experts
2. Marking the blanks $\Rightarrow$ Experts
3. Generation of distractors
4. Choosing the distractors $\Rightarrow$ Experts
5. Evaluation with students $\Rightarrow$ In-class experiment
6. Item Analysis

Experimental settings: Texts

Texts

- 5 articles from zientzia.net

  - Web site that provides current and updated information on Science and Technology in Basque
  - Length of the texts
  - Appropriateness to the students' level

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experimental settings: Blanks

Blanks

- Experts marked between 15 and 20 suitable terms per text
- 94 blanks obtained
  - 69.14% nouns
  - 15.95% verbs

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experimental settings: Heuristics

Heuristics

	Nouns in WordNet		Nouns not in WordNet	Verbs
	Monosemous	**Polysemous**		
Heuristic 1	LSA-based	LSA-based	LSA-based	LSA-based
Heuristic 2	LSA + O + M	LSA + O + M	LSA + O + M	LSA + O + M
Heuristic 3	LSA + Dict.	LSA + Dict.	LSA + Dict.	LSA + Dict.
Heuristic 4	Graph-based	Graph-based	LSA + O + M	LSA + O + M
Heuristic 5	Combination	Combination	LSA + O + M	LSA + O + M
Heuristic 6	Graph-based	Combination	LSA + O + M	LSA + O + M

Table: Heuristics (Legend: O: Ontology; M: Morphology)

Example: Heuristic 6

- Monosemous nouns: Graph-based approach
- Polysemous nouns: Combination approach
- Nouns NOT in WordNet and verbs: LSA & Ontology & Morphology

Experimental settings: Distractors

Distractors

- The expert checked the distractors
  - Correct answer: 1.31%
  - Dubious: 2.96%
  - **Valid distractors: 95.73%**

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Experimental settings: Schools and students

Schools and students

- 18 schools
- 951 students of Obligatory Secondary Education (second grade)
- 30 minutes to read and complete the test (in paper)
- 890 of the students completed the test and their results were used to analyse the items

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

CTT— Measures

- Classical Test Theory
  - Item difficulty
  - Item discrimination
  - Distractor evaluation

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

CTT— Measures

- **Item difficulty**: the proportion of students who can answer the item correctly. The higher the value of difficulty, the easier the item
  - Easy: 90%
  - Difficult: 30%
  - Desired value: 0.5
  - A balance between the number of easy and difficult items
    - In an automatic system, a high number of difficult items is preferable, to a high number of easy items

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

CTT— Measures

- **Item discrimination**: a good item should be able to discriminate students with high ability from those with low ability
  - Results based on low-scoring and high-scoring students
  - A positive value is desirable
  - Problematic: negative values

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

CTT— Measures

- **Distractor evaluation**: to detect poor distractors, the option-by-option responses of high-ability and low-ability students groups will be examined
    - Results based on low-scoring and high-scoring students
    - High selection rate of each distractor
    - Distractors that attract more students from the low-scoring group than from the high-scoring one

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Item difficulty

- The results vary from one text to another (from 0.64 to 0.73)
  - *Bats, Arctic*: Heuristic 6
  - *Earth*: Heuristic 4 $\Rightarrow$ Attain slightly
  - *Continent*: Heuristic 3 better difficulty values

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Item difficulty

- Lowest number of easy items
  - *Bats*, *Arctic*: Heuristic 6
  - *Continent*: Heuristics 6 and 3
  - *Earth*: Heuristics 3 and 1
- No heuristic which create difficult items

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Item discrimination

- All of the heuristic obtained a positive average value
- There is no heuristic which stands out from the rest

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Item discrimination

- No negative discrimination
    - *Earth*: Heuristic 5
    - *Continent*: Heuristics 2, 4 and 6
    - *Bats*: Heuristic 6
    - *Arctic*: Heuristic 2 (one item)

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Distractor analysis

- Overall, Heuristic 6 produced distractors that attracted more students for the *Earth*, *Bats* and *Arctic* texts and Heuristic 4 for the *Continent* text
- Positive discrimination
  - *Bats*, *Arctic*: Heuristic 6
  - *Continent*: Heuristic 4
  - *Earth*: Heuristic 2

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Quantitative analysis

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff.	L > H
Earth (19)	H4: 0.64 (0.23)	H1,3: 1	H5: 3	H5: 0.41 (0.19)	0	H6: 87.72%	H2: 66.67%
Continent (17)	H3: 0.70 (0.23)	H3,6:3	H3: 2	H2 0.26 (0.16)	0	H4: 72.55%	H4: 72.55%
				H4: 0.43 (0.14)	0		
				H6: 0.31 (0.20)	0		
Bats (15)	H6: 0.73 (0.22)	H6: 4	H2,6: 1	H6: 0.41 (0.22)	0	H6: 84.44%	H6: 80.00%
Arctic (14)	H6: 0.66 (0.23)	H6: 1	H6: 2	H2: 0.17 (0.13)	1	H6: 83.33%	H6: 66.67%

Best heuristic

Heuristic 6 is the best overall strategy for producing distractors

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Best heuristic

Best heuristic

- Heuristic 6
  - LSA & ontology & morphology: verbs and nouns NOT in WordNet
  - Graph-based: monosemous nouns in WordNet
  - Combination: polysemous nouns in WordNet
- Heuristic 6 $\Rightarrow$ Fully automatic method

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Real scenario

Real scenario

- Two step item analysis
  - Qualitative analysis: giving the items to experts
  - Quantitative analysis: evaluation of the supervised tests with students
- Heuristic 6 $\Rightarrow$ 10 distractors per item $\Rightarrow$ 3 distractors per item
  - Two step method

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Results

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth (19)							
Two-step	**0.55 (0.28)**	**2 (10.53%)**	**6 (31.58%)**	0.17 (0.17)	3	**87.72**	**71.93**
Fully autom.	0.67 (0.24)	3 (15.79%)	2 (10.53%)	0.16 (0.19)	3	87.72	56.14
Continent (17)							
Two-step	**0.62 (0.18)**	**0 (0.00%)**	**1 (5.88%)**	0.25 (0.16)	**1**	**84.31**	**68.63**
Fully autom.	0.74 (0.18)	3 (17.65%)	0 (0.00%)	0.31 (0.2)	0	66.67	64.71
Bats (15)							
Two-step	0.73 (0.29)	6 (40.00%)	**2 (13.33%)**	0.04 (0.21)	**6**	57.78	46.67
Fully autom.	**0.73 (0.22)**	4 (26.67%)	1 (6.67%)	0.41 (0.22)	0	**84.44**	**80.00**

Item difficulty

- The overal item difficulty is closer to the desired value of 0.5 (*Bats* exception)
- The number of easy items decreases
- The number of difficult items increases

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Results

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth (19)							
Two-step	**0.55 (0.28)**	**2** (10.53%)	**6** (31.58%)	0.17 (0.17)	3	**87.72**	**71.93**
Fully autom.	0.67 (0.24)	3 (15.79%)	2 (10.53%)	0.16 (0.19)	3	87.72	56.14
Continent (17)							
Two-step	**0.62 (0.18)**	**0** (0.00%)	**1** (5.88%)	0.25 (0.16)	**1**	**84.31**	**68.63**
Fully autom.	0.74 (0.18)	3 (17.65%)	0 (0.00%)	0.31 (0.2)	0	66.67	64.71
Bats (15)							
Two-step	0.73 (0.29)	6 (40.00%)	**2** (13.33%)	0.04 (0.21)	**6**	57.78	46.67
Fully autom.	**0.73 (0.22)**	**4** (26.67%)	1 (6.67%)	0.41 (0.22)	0	**84.44**	**80.00**

Item discrimination

- *Earth* and *Continent*: no difference in the results
- *Bats*: the number of items that discriminates negatively in the *Bats* texts increases substantially

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Results

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth (19)							
Two-step	**0.55 (0.28)**	**2 (10.53%)**	**6 (31.58%)**	0.17 (0.17)	3	**87.72**	**71.93**
Fully autom.	0.67 (0.24)	3 (15.79%)	2 (10.53%)	0.16 (0.19)	3	87.72	56.14
Continent (17)							
Two-step	**0.62 (0.18)**	**0 (0.00%)**	**1 (5.88%)**	0.25 (0.16)	**1**	**84.31**	**68.63**
Fully autom.	0.74 (0.18)	3 (17.65%)	0 (0.00%)	0.31 (0.2)	0	66.67	64.71
Bats (15)							
Two-step	0.73 (0.29)	6 (40.00%)	**2 (13.33%)**	0.04 (0.21)	6	57.78	46.67
Fully autom.	**0.73 (0.22)**	**4 (26.67%)**	1 (6.67%)	0.41 (0.22)	0	**84.44**	**80.00**

Distractor analysis

- The number of selected distractors and positive discrimination increases
  - Exception: *Bats*

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Discussion

Discussion

- The results applying a two step method are better
- Even for experts the generation of distractors is a difficult task

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Generation of distractors by hand

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth	0.46 (0.26)	0 (0.00%)	7 (36.84%)	0.14 (0.21)	**4**	92.98	68.42
Continent	0.66 (0.23)	0 (0.00%)	1 (5.88%)	0.13 (0.21)	**4**	74.51	56.86
Bats	0.60 (0.25)	1 (6.67%)	1 (6.67%)	0.05 (0.15)	**5**	68.89	46.67
Arctic	0.70 (0.21)	3 (21.43%)	1 (7.14%)	0.06 (0.15)	**6**	61.90	54.76

Item difficulty

- Overall, the manual method obtains better results
- The expert tends to create more balanced tests in terms of difficulty
- The two step method obtains similar results in terms of item difficulty

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Generation of distractors by hand

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth	0.46 (0.26)	0 (0.00%)	7 (36.84%)	0.14 (0.21)	**4**	92.98	68.42
Continent	0.66 (0.23)	0 (0.00%)	1 (5.88%)	0.13 (0.21)	**4**	74.51	56.86
Bats	0.60 (0.25)	1 (6.67%)	1 (6.67%)	0.05 (0.15)	**5**	68.89	46.67
Arctic	0.70 (0.21)	3 (21.43%)	1 (7.14%)	0.06 (0.15)	**6**	61.90	54.76

Item discrimination

- More items with negative discrimination
- The expert tends to create more difficult items so that the distractors would be more attractive to the entire group of students

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Generation of distractors by hand

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth	0.46 (0.26)	0 (0.00%)	7 (36.84%)	0.14 (0.21)	**4**	92.98	68.42
Continent	0.66 (0.23)	0 (0.00%)	1 (5.88%)	0.13 (0.21)	**4**	74.51	56.86
Bats	0.60 (0.25)	1 (6.67%)	1 (6.67%)	0.05 (0.15)	**5**	68.89	46.67
Arctic	0.70 (0.21)	3 (21.43%)	1 (7.14%)	0.06 (0.15)	**6**	61.90	54.76

Distractor analysis

- The manual method obtains better results as regards the percentage of selected distractors

- The manual method tends to discriminate negatively more

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Generation of distractors by hand

	Item difficulty			Item discrimination		Distractor analysis	
	Overall	Easy	Difficult	Overall	Neg.	Diff. (%)	L > H (%)
Earth	0.46 (0.26)	0 (0.00%)	7 (36.84%)	0.14 (0.21)	**4**	92.98	68.42
Continent	0.66 (0.23)	0 (0.00%)	1 (5.88%)	0.13 (0.21)	**4**	74.51	56.86
Bats	0.60 (0.25)	1 (6.67%)	1 (6.67%)	0.05 (0.15)	**5**	68.89	46.67
Arctic	0.70 (0.21)	3 (21.43%)	1 (7.14%)	0.06 (0.15)	**6**	61.90	54.76

Discussion

- Overall results of the expert are better
- The expert is unable to generate distractors of the same quality for different texts

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Replacement of the keys

Manual method

- Item difficulty
  - Average between 0.6 and 0.7
  - One text: 0.46
- The number of occurrences of some keys high
  - *Continent*: 64% of the keys more than once; average frequency 7.81

Hypothesis

The repetition of the key could help students selecting the correct answer

Experiment

- The keys of the items replaced with a synonym

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Replacement of the keys

Results — Item difficulty

	Item difficulty			
	Manually generated		Automatically generated	
	Original	Synonyms	Original	Synonyms
Item 2	0.66	**0.16**	0.84	**0.48**
Item 3	0.84	**0.55**	0.94	**0.74**
Item 4	0.84	**0.39**	0.75	**0.56**
Item 5	0.81	**0.16**	0.56	**0.44**
Item 7	0.81	**0.39**	0.84	**0.52**
Item 8	0.72	**0.26**	0.78	**0.37**
Item 16	0.59	**0.42**	**0.81**	0.89
Item 17	**0.75**	0.87	**0.44**	0.67

- The manually generated distractors: increases the difficulty of 7 out of 8
- Automatically generated distractors: increases the difficulty of 6 out of 8
- Items with no synonyms: variation of 0.1
- Overall
  - Manual method: 0.66 to 0.48
  - Automatic method: 0.74 to 0.65

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

Replacement of the keys

Discussion

- The different occurrences of the key within the texts help students to complete the tests

- Our system should consider the option of replacing the key with a synonym

Introduction
Architecture
Model
Experiments
Conclusions and future work

Question generation
Declension and verb tests
Determiner tests
MCQs for English verb tests
MCQs for Basque science tests

To sum up

To sum up

- Six heuristics which depend on the PoS and polysemy of nouns
- Fully automatic method — Best heuristic
  - LSA & ontology & morphology: verbs and nouns NOT in WordNet
  - Graph-based: monosemous nouns in WordNet
  - Combination: polysemous nouns in WordNet
- Similar results
  - Two step method and manual method

Two objectives

- A helping tool for experts
- Automatic generation of items of a certain standard

Outline

1. Introduction

2. Architecture

3. Model

4. Experiments

5. **Conclusions and future work**

Conclusions

Motivation

ICTs are widely used in different scenarios as media and methodologies

Contribution

- ICT as an approach to help in the learning process of certain subjects
- The analysis of various available NLP tools and corpora has demonstrated that it is possible to implement a system that helps experts and teachers in the creation of didactic material

Conclusions

ArikIturri

- We have designed and implemented a system called **ArikIturri** that, based on NLP and corpora, is able to produce items of a certain standard
- ArikIturri is a multilingual system, and different question types have been tested in several scenarios

Question model

- The representation of the items as well as the information relating to their generation process is carried out by means of a question model
- This structured representation allows the importation and exportation of the items into independent applications

Conclusions

Experiments

1. To prove the viability of the system designed to implement a complete automatic process to generate items
2. To apply different methods in the generation of distractors
   - Morphology, syntax, semantic
3. To modify some components of the source sentences when creating the stems

Results

- Experts' opinions and students' answers
  - A qualitative analysis based on experts knowledge gave us a way of measuring the correctness of the automatically generated questions
  - The quantitative analysis based on students' responses ensured the quality of the items

Future work

Topic selection

- To automatise the process of detecting the meaningful terms in the text

Question generation challenge

- Automatic generation of questions
- Generation of relevant questions
- Reading comprehension task

Eskerrik asko